# Thai Food Safety Document Searching System by Ontology

Ampaphun Wijasika, and Anongnart Srivihok

***Abstract*—**Since Thailand is an agricultural country, most of Thai research papers are about agricultural and food safety. Thai researchers produce so many food safety research documents which the numbers are growing up each year. The searching system for food safety documents is needed. The objective of this paper is to introduce the development of searching system for Thai food safety documents by using ontology. This system can help users who want to search for food safety documents and get the relevant information. Also, food safety ontology was created by the collection of terms or keywords from document titles and an abstracts and using Hozo ontology editor. This ontology is a significant component of document searching system.

***Keywords*—** Food safety, ontology, information searching.

## I. Introduction

FOOD is one of the critical factors for human being since everyone needs food for energy and growing. Further, the safety of food to be taken is more significant. Thailand, as a leading food exporter country, is emphasizing in producing food to meet the safety standard. Then, Food Safety is a hot topic for researchers to study and publish at present.

For helping farmers and food industry to work and produce high quality of agriculture products and decreasing cost, many Thai researchers have studied and published many research documents in food safety area. In 2012, there are more than 5900 publications were produced and the numbers are increasing about 6% every year (surveyed by Library of Chaingmai University, Thailand. April, 19 2013). It implied that the amount of research documents is growing each year. These bring about problems in collecting, storing, cataloging, linking, and retrieving both data and information. At present, the system which provides the facility in searching and retrieving appropriate information is needed.

We propose the Thai food safety document searching system based on ontology (FSO) which no one builds the system with this domain in Thailand. We divided the system into two parts. First, we build the food safety ontology that keeps the keywords or terms from document title in hierarchical (ontology) structure and we build the document searching system based on ontology.

Ampaphun Wijasika is with the Department of Computer Science, Faculty of Science, Kasetsart University, Bangkok, Thailand 10900. E-mail: g5314401321@ku.ac.th

Anongnart Srivihok, is with the Department of Computer Science, Faculty of Science, Kasetsart University, Bangkok, Thailand 10900. E-mail: fsciang@ku.ac.th

The remainder of this paper is organized as follow. First, the related works is presented in section 2. In section 3, we described a methodology and framework. In section 4, we described experiments and discussion. Finally, conclusion and future work in the last section.

## II. Relate works

### A. What is an Ontology

Ontology [1] is represented by the conceptualization of collected attributes or terms within the same domain. Moreover, the attributes or terms are related between each other and dimensions. To build an ontology, we have to scope the domain that we are interested in. Then we can collect the knowledge or terms within our specific domain to generate ontology.

### B. An ontology with Information system

Furthermore, an ontology can be used to represent the specific terms with their relationships. Accordingly an ontology can be applied with various kinds of techniques to build information systems, such as data mining, and knowledge management. For example; the healthy recommendation system and expert system for generic name drug in tablet dosage form [2].

### C. Document Modification

Hozo ontology editor [3] is a freeware which developed by Osaka University, Japan since 2007. Hozo is a tool using for generating an ontology and it also provides graphical interface. Hozo syntax is an Extensible Markup Language (XML). Moreover Hozo can be exported in Web Ontology Language (OWL), Resource Description Framework (RDF) and Resource Description Framework Schema (RDFS) which all are W3C standards.

### D. Related researches

In Ontology, the helpfulness of hierarchy information is making us understand and see a relation between terms in the same domain. In addition, we can apply the ontology to share or reuse the knowledge.

In Thailand, where agriculture industry is the traditional theme of this country, Heeptaisong and Srivihok[4] proposed a soil ontology that represented knowledge about soil from different kinds of sources such as, electronic documents, papers or books by Hozo ontology editor. Then, they built the system for soil knowledge searching. In addition, Tongjing[5] proposed an ontology for integration of Thai herbal knowledge by using Protege ontology editor. The system

provided information about various kinds of Thai herbs properties that had different effects with the disease.

In different domain, Sooksom, Booranaratch, Subnithi and Netisopakul[6] proposed the personalized food recommendation system. They generated a food and nutrition ontology that combined with the rule of personal profile to provide recommendations based on nutrition requirements. Chalortham and Leesawat[7] proposed the pharmaceutical ontology combined with the skilful of pharmacists which represented by rules for generic equivalent tablet production and herb tablet production. There were 61 classes in pharmaceutical ontology and 5 relationships. Both researches used Hozo ontology editor to generate ontology models.

In addition, ontology can be driving to specific knowledge such as Shivakumar and Porkodi[8] proposed three biological ontologies (1) cellular components, (2) molecular functions, and (3) biological processes. Then they used association rule mining to uncover the relationships between different gene products in different genes. Accordingly, Yang, Du and Hu[9] also proposed the food safety ontology reasoning application. In different way from our work, they created reasoning rules for collecting food safety knowledge based on Jena API which included three parts (1) creation part, (2) modification part and (3) query for retrieve related documents.

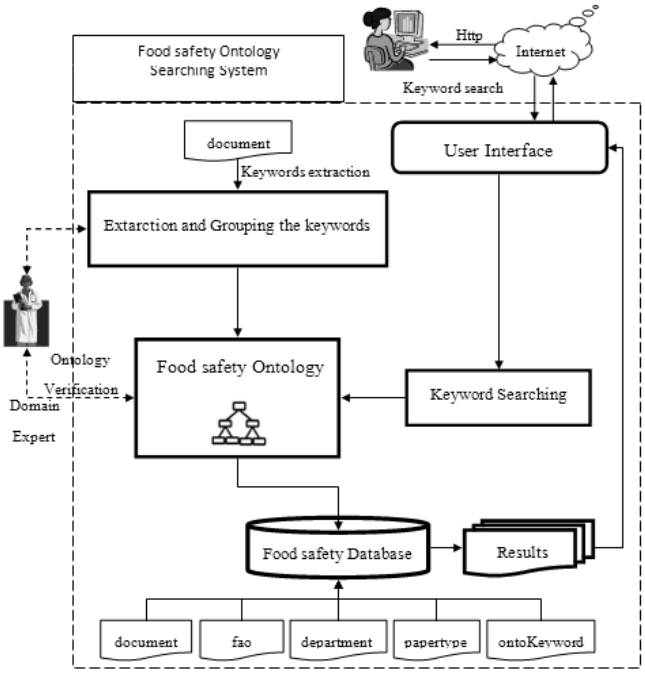


Fig. 1 Food safety document searching system by ontology (FSO) Framework

## III. Methodology and Framework

The study framework is shown in Fig.1. We divided the system into two steps. First, we built the food safety ontology. Second, we built the document searching system based on the built ontology.

We used a Thai food safety document database which including 1,642 documents of master thesis of Thai Universities from 2004 – 2006. There are four tables including document table, department table, fao table and papertype table.

The Steps of Food ontology searching system are depicted as follows:

Step 1: Building food safety ontology as shown in Fig.2;

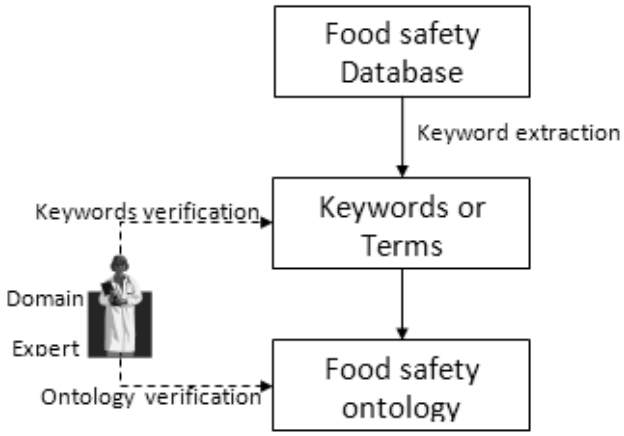


Fig.2 Food safety ontology framework

### A. Food safety Database

Document table included the details of documents there are 14 attributes such as document number, title, author name and key word. Department table included 5 attributes such as, department_id, and department_name of the authors. Fao table included subdomain of document subject according to Food and Agriculture Organization. It was divided into 14 groups such as agriculture, food forestry, fishery and others. There were 3 attributes such as, fao_id, and fao_name. Papertype table included 2 attributes papertype_id and papertype_name.

### B. Extraction and grouping the keywords

First of all, 424 keywords were extracted from document titles. Then, all keywords were segmented into two big groups: food and food safety and put keywords in each group by relation (levels). The results from grouping are showed in Fig. 3. In this process, a food safety expert from a public university checked the correctness of this ontology.

### C. Food safety ontology

After extracting and grouping all keywords, we brought all keywords to build the food safety ontology by using Hozo editor. As well, we can also obtain OWL structure of food safety ontology that generated by Hozo ontology editor, which is depicted in Fig. 4. This figure includes keyword "อาหาร" (food) in level 1, keyword "สัตว์" (animal) in level 2 and keyword "พืช" (plant) in level 2. Then, Food Safety ontology represented by hierarchical structure in Fig. 5.

| | Level 1 | Level 2 | Level 3 | Level 4 | Level 5 | Level 6 |
|---|---|---|---|---|---|---|
| หมวดหมู่ของคำสำคัญ (Food safety keywords) | อาหาร (Food) | สัตว์ (Animal) | สัตว์เลี้ยง (Farm animal) | สัตว์น้ำ (Aquatic animals) | ปลา (Fishes) | ... |
| | | | | | กุ้ง (Shrimps) | ... |
| | | | | | หอย (Shells) | ... |
| | | | | สัตว์บก (Land animals) | ... | ... |
| | | พืช (Plant) | พืชไร่ (Farm plant) | พืชเศรษฐกิจ (Industrial Plant) | ยางพารา (Rubber tree) | ... |
| | | | | | ปาล์ม (Palm tree) | ... |
| | | | | | สัก (Teak) | ... |
| | | | พืชสวน (Horticultures) | ผัก (Vegetables) | ผักใบก้าน (Leaf Veg.) | ... |
| | | | | | ผักหัว (Under soil Veg.) | ... |
| | | | | | ผักผล (Fruit Veg.) | ... |
| | | | ธัญพืช(Grains) | ถั่ว (Beans) | ... | ... |
| | | | | ข้าว (Rice) | ข้าวหอม (Jasmine rice) | ข้าวหอมมะลิ (Thai Jasmine) |
| | | | | | | ข้าวหอมนิล (Thai black Jasmine rice) |
| | | | | | ข้าวเจ้า (White rice) | ... |
| | | | | | ข้าวเหนียว (Glutinous rice) | ... |
| | | | | | ข้าวสาลี (Wheat) | ... |
| | | | | งา (Sesamum) | ... | ... |

Fig. 3 The Food Safety keywords that are shown in a group in order by level.

```
<owl:Class rdf:ID="Any">
  <rdfs:label>Any</rdfs:label>
</owl:Class>
<owl:Class rdf:ID="อาหาร">  (Food = Level 1)
  <rdfs:label>อาหาร</rdfs:label>
  <rdfs:subClassOf rdf:resource="#Any" />
</owl:Class>
<owl:Class rdf:ID="สัตว์">  (Animal = Level 2)
  <rdfs:label>สัตว์</rdfs:label>
  <rdfs:subClassOf rdf:resource="#อาหาร" />  (Food = Level 1)
</owl:Class>
<owl:Class rdf:ID="พืช">  (Plant = Level 2)
  <rdfs:label>พืช</rdfs:label>
  <rdfs:subClassOf rdf:resource="#อาหาร" />  (Food = Level 1)
</owl:Class>
```

Fig.4 Food safety Ontology is depicted in OWL code

*D.Food safety ontology verification*

After food safety ontology was created by using Hozo ontology editor. A food safety expert had reviewed this ontology for the correctness of the structure and relationships of nodes.

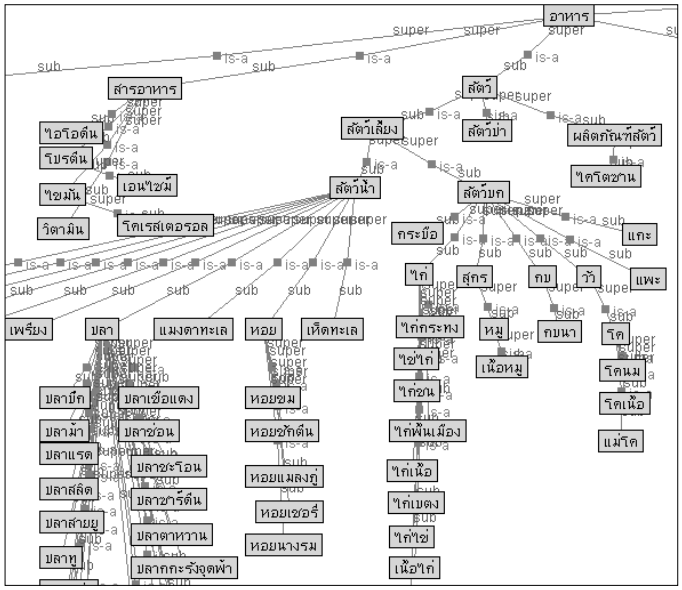


Fig. 5 Food safety ontology by Hozo ontology editor

Step 2: Building searching system based on Food safety;

| KeywordID | KeywordName | | Parent_ID |
|---|---|---|---|
| 0100000000 | อาหาร | (Food) | 0000000000 |
| 0110000000 | น้ำ | (Water) | 0100000000 |
| 0111000000 | น้ำดื่ม | (Drinking water) | 0110000000 |
| 0112000000 | น้ำแข็ง | (Ice) | 0110000000 |
| 0120000000 | สารอาหาร | (Nutrition) | 0100000000 |
| 0121000000 | ไอโอดีน | (Iodine) | 0120000000 |
| 0122000000 | ไขมัน | (Fat) | 0120000000 |
| 0122100000 | โคเรสเตอรอล | (Cholesterol) | 0122000000 |
| 0123000000 | โปรตีน | (Protein) | 0120000000 |
| 0123100000 | เอนไซม์ | (Enzyme) | 0123000000 |
| 0124000000 | วิตามิน | (Vitamin) | 0120000000 |
| 0130000000 | สัตว์ | (Animal) | 0100000000 |
| 0131000000 | สัตว์เลี้ยง | (Farm Animal) | 0130000000 |
| 0131100000 | สัตว์น้ำ | (Aquatic Animal) | 0131000000 |
| 0131110000 | ไดอะตอม | (Diatom) | 0131100000 |
| 0131111000 | เห็ดทะเล | (Mushroom Anemone) | 0131100000 |
| 0131111100 | โพรโทซัว | (Protozoa) | 0131100000 |

Fig. 6 The ontokeyword table

*A. The ontokeyword table*

Before FSO is built, we have to create the table which keeps the relationship between each keyword. The ontokeyword table is containing three attributes which shown as depicted in Fig. 6. KeywordID is the field which stores all keyword's id, KeywordName is the field that places all keyword names and lastly, Parent_ID is the parent id of

related keywords.

We get the information from the hierarchy level of each keyword which generated by Hozo, as depicted in Fig.7. An example, Cholesterol is lower level of Fat then Parent_ID of Cholesterol is Fat's ID, Fat is under the level of Nutrition then Parent_ID of Fat is Nutrition's ID respectively.

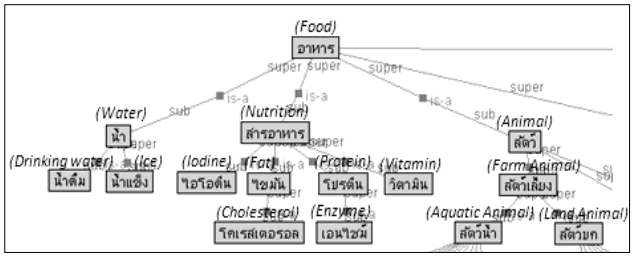


Fig. 7 Some part of hierarchy level of food safety keyword

*B. Food safety document searching system based on ontology*

We are building the system for searching. Which can get the input (keyword) from user as depicted in Fig. 8.

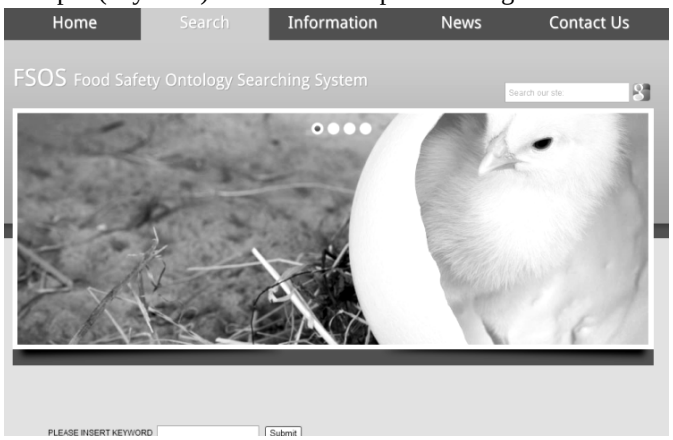


Fig.8 The user interface of Thai Food Safety Searching System (FSO).

## IV. Experiment and Discussions

This section is about the experimental and result discussions. In FSO system, we put the target keyword into the text field and press submit button, the searching system compared a target keyword from user with all keywords in ontoKeyword table. After comparing and matching keywords have been finished, the system retrieves all keywords that are related to the target keyword by reference with KeywordID and Parent_ID. Next, all keywords which relate to the target keyword are displayed. In Fig.9, we insert the target keyword as "ธัญพืช" (grain). The searching system shows all keywords that relate to "ธัญพืช" (grain). From the example, FSO lists 5 keywords that related to "ธัญพืช" (grain). There are "งา" (sesame), "มอลต์" (malt), "ลูกเดือย" (job's tear), "ถั่ว" (bean) and "ข้าว" (rice).

There are 368 documents containing five keywords: "ธัญพืช" (grain). There are "งา" (sesame), "มอลต์" (malt), "ลูกเดือย" (job's tear), "ถั่ว" (bean) and "ข้าว" (rice). Moreover, the system also provides the more information of their descriptions to user. Such as, document title, author name, source of publication and published year.

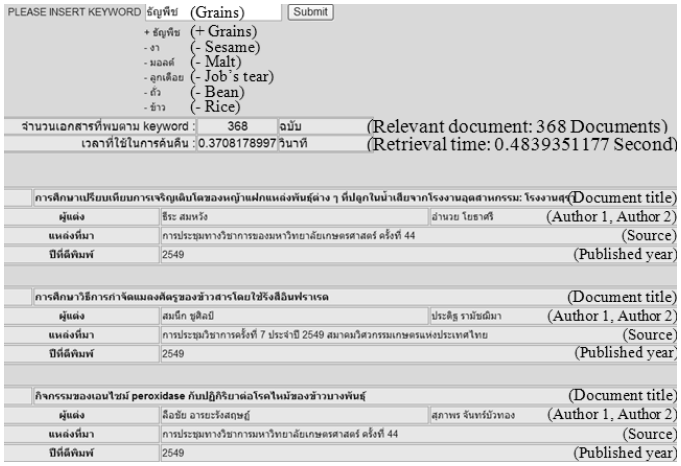


Fig.9 The example information that searching based on ontology

We evaluated document retrieval results by using precision and recall[10]. The index for evaluation is showed by equation (1) and equation (2), respectively.

$$Precision(P) = \frac{P(A \cap B)}{\mathrm{P(B)}} = \frac{relevant\ transactions}{\text{all relevant transactions}} \quad (1)$$

$$Recall(R) = \frac{\mathrm{P(A \cap B)}}{\mathrm{P(A)}} = \frac{relevant\ transactions}{\text{relevant transactions in system}} \quad (2)$$

Where, A is the set of transactions in the system and B is the relevant set of transactions querying in the retrieval system.

TABLE I
Precision and Recall

| *Retrieval system* | *Precision* | *Recall* |
|---|---|---|
| Traditional Food safety Searching system | 42.457% | 64.426% |
| Food safety ontology Searching system (FSO) | 71.602% | 80.396% |

We test our searching system with the traditional searching system (no ontology applied) with the example in following.

Example 1: We use a keyword "ปู" (Crab). There are 49 documents which include "ปู" (Crab) in document titles. The tradition Food safety Searching system without ontology retrieves 89 documents while FSO showed 12 documents from all 14 relevant documents.

Example 2: We use a keyword "เกลือ" (Salt). There are 13 documents which include "เกลือ" (Salt) in document title. The tradition Food safety Searching system without ontology retrieves 46 documents while FSO retrieved 11 documents from all 13 relevant documents.

Example 3: We use a keyword "งา" (Sesame). There are 89 documents which include "งา" (Sesame) in document titles. The tradition Food safety Searching system without ontology

retrieves 250 documents while FSO showed 30 documents from all 61 relevant documents.

Example 4: We use a keyword "แกะ" (Sheep). There are 9 documents which have "แกะ" (Sheep) in document title. The tradition Food safety Searching system without ontology retrieves 13 documents while FSO showed 1 document from 1 relevant document.

Example 5: We use a keyword "กบ" (Frog). There are 12 documents which include "กบ" (Frog) in document titles. The tradition Food safety Searching system without ontology retrieves 102 documents while FSO showed 6 documents from all 20 relevant documents.

Example 6: We use a keyword "โค" (Cow). There are 37 documents which have "โค" (Cow) in document title. The tradition Food safety Searching system without ontology retrieves 119 documents while FSO showed 134 documents from all 185 relevant documents.

Example 7: We use a keyword "หมาก" (Betel nut). There are 9 documents which have "หมาก" (Betel nut) in document title. The tradition Food safety Searching system without ontology retrieves 16 documents while FSO showed 4 documents from all 5 relevant documents.

## V. Conclusion and Future work

The aim of this research is to develop a food safety searching system (FSO) for research documents. The food safety ontology was built from keywords of published documents in Thailand. The FSO system is based on food safety ontology hierarchical structure; this system can provided the relevant documents which related to given keywords search with 71% precision and 81% recall.

For future work, we will apply the association rule mining to document keywords to find the relationships between them. This can be applied to the searching system for retrieval expanding to get more related documents.

## References


[1] N. Guarino, "Formal Ontology and Information Systems," Proceedings of FOIS, pp. 3-15, Italy: IOS Press, 1998.

[2] N. Chalortham, P. Leesawat, M. Booranaratch and T. Subnithi, "An expert system for generic name drug in tablet" *NETTEC Technical Journal, NECTEC-ACE2009 Special Edition.* [September 2009].

[3] R. Mizoguchi, E. Sunagawa, K. Kozaki and Y. Kitamura (April 2007). The model of roles within an ontology development tool: Hozo. IOS Press.[Online]. pp. 159-179. Available:http://iospress.metapress.com/content/w67u25284x0l206v/

[4] T. Heeptaisong and A. Srivihok, "Ontology Development for searching soil knowledge," *The 9th International conference on e-Business*, Thailand:iNCEB, 2010.

[5] P. Tongjing and R. Peechawanich, "Ontology for semantic integration of Thai herbal knowledge," M.S. thesis, Dept. Computer Science, Thammasat Univ., Bangkok, Thailand, 2010.

[6] N. Sooksom, M. Booranaratch, T. Subnithi and P. Netisopakul, "Ontology Development for Personalization Food and Nutrition Recommender System," *NECTEC Technical Journal, NECTEC-ACE2010 Special Edition.*

[7] N. Chalortham, P. Leesawat, "An Expert System for Tablet Production of Herbal and Generic Drug" Ph.D. thesis, Dept. Pharmacy, Chiang Mai Univ., Chiang Mai, Thailand, 2010.

[8] B.L.Shivakumar and R. Porkodi. Finding relationship among gene ontology terms in biological documents using Association Rule mining and GO annotation. *IRACST International Journal of Computer Science and Information Technology & Security (IJCSITS).* June 2012.

[9] Y. Yang, J. Du and J. Hu, "Study on food safety otology reasoning application based on jena" Proceedings of ICCTA, pp.727-731, 2011

[10] D.M.W. (February 2011). Evaluation: from precision, recall and f-measure to ROC, informedness, markedness& correlation. *Journal of Machine Learning Technologies.*



**Ampaphun Wijasika**, received the Bachelors Degree (B.Sc.) in Computer Science from King Mongkut's Institute of Technology Ladkrabang, Bangkok, Thailand in the year 2010, and currently she is pursuing her master degree Computer Science, Department of Science at Kasetsart University, Bangkok, Thailand, in the area of Ontology and Data mining.

**Assoc. Prof. Anongnart Srivihok**, Department of Science at Kasetsart University, Bangkok, Thailand.